Freeform Search

Database: US Pre-Grant Publication Full-Text Database
US Patents Full-Text Database
US OCR Full-Text Database
EPO Abstracts Database
JPO Abstracts Database
Derwent World Patents Index
IBM Technical Disclosure Bulletins

Term:

Display: 10 **Documents in Display Format:** - **Starting with Number** 1

Generate: ○ **Hit List** ◉ **Hit Count** ○ **Side by Side** ○ **Image**

Search Clear Interrupt

Search History

DATE: Wednesday, August 01, 2007 **Purge Queries** Printable Copy Create Case

Set Name side by side	Query	Hit Count	Set Name result set
	DB=PGPB,USPT,USOC,EPAB,JPAB,DWPI,TDBD; PLUR=YES; OP=OR		
L93	l57 and l84	39	L93
	DB=DWPI,TDBD; PLUR=YES; OP=OR		
L92	l57 and l84	0	L92
	DB=EPAB,JPAB; PLUR=YES; OP=OR		
L91	l57 and l84	0	L91
	DB=PGPB,USPT,USOC; PLUR=YES; OP=OR		
L90	l57 and l84	39	L90
	DB=USPT; PLUR=YES; OP=OR		
L89	'6105020'.pn.	1	L89
L88	'6397204'.pn.	1	L88
L87	'6516310'.pn.	1	L87
L86	'7010516'.pn.	1	L86
	DB=PGPB,USPT,USOC,EPAB,JPAB,DWPI,TDBD; PLUR=YES; OP=OR		
L85	L84 and (customer with profile with groups or customer near profile near groups or customer adj profile adj groups)	8	L85

L84	L83 and (fact with table or fact near table or fact adj table)	81	L84
L83	l38 and (dimension with table or dimension near table or dimension adj table)	86	L83
	DB=USPT; PLUR=YES; OP=OR		
L82	(5544298 | 5088052 | 5471611 | 5414838 | 5537590 | 5455945 | 5404506)![PN]	7	L82
L81	("5832496")[PN]	1	L81
	DB=PGPB,USPT,USOC,EPAB,JPAB,DWPI,TDBD; PLUR=YES; OP=OR		
L80	5832496.pn.	2	L80
L79	5832498.pn.	2	L79
L78	5615109.pn.	2	L78
	DB=USPT; PLUR=YES; OP=OR		
L77	'6438537'.pn.	1	L77
L76	'6434557'.pn.	1	L76
L75	'6374263'.pn.	1	L75
L74	'6385201'.pn.	1	L74
L73	'6477525'.pn.	1	L73
L72	'6505205'.pn.	1	L72
L71	'6505205'.pn.	1	L71
L70	'6487546'.pn.	1	L70
L69	'6484179'.pn.	1	L69
L68	'6480836'.pn.	1	L68
L67	'6173310'.pn.	1	L67
L66	'6173310'.pn.	1	L66
L65	'6151601'.pn.	1	L65
L64	'6151601'.pn.	1	L64
L63	'6141655'.pn.	1	L63
	DB=PGPB,USPT,USOC,EPAB,JPAB,DWPI,TDBD; PLUR=YES; OP=OR		
L62	L57 and 717.clas.	2	L62
L61	L57 and 706.clas.	2	L61
L60	L57 and 705.clas.	19	L60
L59	L57 and 707.clas.	146	L59
L58	L57 not @py>1999	0	L58
L57	L56 and (metadata or metadata with model or meta-data near model or meta-data adj model)	168	L57
L56	(star with schema or star near schema or star adj schema or "reverse star schema" or reverse with star with schema or reverse near star near schema or snowflake with schema or snowflake near schema or snowflake adj schema) and ("data warehouse" or datamart)	324	L56
	DB=USPT; PLUR=YES; OP=OR		
L55	'5995958'.pn.	1	L55
L54	'5806060'.pn.	1	L54
L53	'5386556'.pn.	1	L53
L52	'5854746'.pn.	1	L52

L51	'5873096'.pn.	1	L51
L50	'5978788'.pn.	1	L50
L49	'6032158'.pn.	1	L49
L48	'6032158'.pn.	1	L48
L47	'6668253'.pn.	1	L47
L46	'6295551'.pn.	1	L46
L45	'6606708'.pn.	1	L45
L44	'6226752'.pn.	1	L44
L43	'6715080'.pn.	1	L43
L42	'6715080'.pn.	1	L42
L41	'6782425'.pn.	1	L41
	DB=PGPB,USPT,USOC,EPAB,JPAB,DWPI,TDBD; PLUR=YES; OP=OR		
L40	L39 and (populat$ and datamart or populat$ and data near2 mart or populat$ adj data adj mart)	24	L40
L39	L38 and (business and database or business and data near2 base or business adj database or business same database)	95	L39
	DB=USPT; PLUR=YES; OP=OR		
L38	("star schema" or "reverse star schema") and (datawarehouse or data with warehouse or data with mart or datamart)	109	L38
L37	("5799286")[URPN]	74	L37
L36	("5799286")[PN]	1	L36
	DB=PGPB,USPT,USOC,EPAB,JPAB,DWPI,TDBD; PLUR=YES; OP=OR		
L35	5799286.pn.	2	L35
	DB=USPT; PLUR=YES; OP=OR		
L34	("6212524")[URPN]	38	L34
L33	("6212524")[URPN]	38	L33
	DB=PGPB,USPT,USOC,EPAB,JPAB,DWPI,TDBD; PLUR=YES; OP=OR		
L32	6212524.pn.	2	L32
	DB=USPT; PLUR=YES; OP=OR		
L31	("6377993")[URPN]	76	L31
L30	(5041972 | 5481542 | 5999973 | 5958016 | 5245533 | 5790797 | 6065002 | 5844896 | 5920542 | 5721913 | 6078891 | 5790780 | 5530744 | 5825769 | 5991733 | 5819225 | 5787412 | 5835084 | 5548726 | 6011844 | 5327486 | 5877759 | 5781550 | 5727129 | 5610915 | 5884032 | 5602918 | 5742905 | 5819271 | 6058381 | 6049789 | 6023762 | 4893248 | 5796393 | 5764756 | 5907681 | 5812533 | 5742768 | 5963925 | 5826269 | 4817050 | 5930764 | 5696906 | 5778377 | 5870558 | 5315093 | 5526257 | 5826029 | 5325290 | 5490060 | 5815080 | 5793964 | 5721908 | 5778178 | 6041325 | 6032132 | 5991806 | 5742763 | 5369571 | 6119109 | 5692181 | 6014647 | 5563805 | 5768501 | 6115693 | 5923016 | 6115737 | 6073105 | 5475836 | 5452446 | 5689645 | 6145001 | 5805803 | 5862325 | 5621727 | 6161128 | 6078924 | 6065059 | 6137869 | 5848396 | 5815665 | 6134584 | 5649182 | 5742762 | 4345315 | 6212506 | 5491779 | 5537611 | 5228076 | 5966695 | 6049602 | 5699403 | 5852810 | 6091808 | 6014702 | 5961602 | 6085171 | 4160129 | 5781632 | 5892900 | 5845067 | 5909682 | 5799154 | 5974396 | 5974441 | 5566351 | 5671354 | 6115458 | 6094655 | 5745754 | 5734831 | 5812654 |	156	L30

	5999972 | 6253239 | 5802320 | 5131020 | 5793762 | 6044144 | 5790809 | 4972504 | 5852812 | 5790789 | 5734709 | 6031904 | 5915001 | 5623601 | 5287270 | 5754830 | 5812750 | 5551025 | 6085190 | 6212558 | 6032184 | 5774660 | 5909679 | 6131095 | 5666481 | 5692030 | 6064667 | 5930804 | 5970467 | 6128624 | 6073241 | 5706502 | 5982864 | 5483596 | 6115040 | 5075771 | 5630066 | 5699528 | 5787160 | 5708780 | 5850517 | 5982891 | 5285494 | 5845267)![PN]		
L29	("6377993")[PN]	1	L29
	DB=PGPB,USPT,USOC,EPAB,JPAB,DWPI,TDBD; PLUR=YES; OP=OR		
L28	6377993.pn.	2	L28
L27	717/105	553	L27
L26	717/104	891	L26
L25	717/102	271	L25
L24	706/52	645	L24
L23	705/44	1408	L23
L22	705/39	2336	L22
L21	705/35	3173	L21
L20	705/30	1335	L20
L19	705/28	2368	L19
L18	705/26	7782	L18
L17	705/16	1225	L17
L16	705/14	5636	L16
L15	705/7	3185	L15
L14	705/5	1193	L14
L13	705/1	7432	L13
L12	707/206	1595	L12
L11	707/201	4030	L11
L10	707/200	6242	L10
L9	707/104.1	8830	L9
L8	707/101	6471	L8
L7	707/100	10569	L7
L6	707/10	15344	L6
L5	707/1	9888	L5
L4	717.clas.	14525	L4
L3	706.clas.	8287	L3
L2	705.clas.	52424	L2
L1	707.clas.	57191	L1

END OF SEARCH HISTORY

IEEE Xplore® RELEASE 2.3

Search Result - Print Format

< Back t

Key: IEEE JNL = IEEE Journal or Magazine, IEE JNL = IEE Journal or Magazine, IEEE CNF = IEEE Conference, II CNF = IEE Conference, IEEE STD = IEEE Standard

1. **Dynamic multi-dimensional models for text warehouses**
   Bleyberg, M.Z.; Ganesh, K.;
   Systems, Man, and Cybernetics, 2000 IEEE International Conference on
   Volume 3, 8-11 Oct. 2000 Page(s):2045 - 2050 vol.3
   IEEE CNF

2. **Data warehouse design for manufacturing execution systems**
   Kai-Ying Chen; Teh-Chang Wu;
   Mechatronics, 2005. ICM '05. IEEE International Conference on
   10-12 July 2005 Page(s):751 - 756
   IEEE CNF

3. **TSMC turnkey data mart**
   Sung-Ting Hsieh, D.; Cheng-Chin Feng, E.; Wei-Ling Liu; I-Chieh Chung;
   Semiconductor Manufacturing Technology Workshop, 2002
   10-11 Dec. 2002 Page(s):267 - 270
   IEEE CNF

4. **A design and practical use of spatial data warehouse**
   Ji-man Park; Chul-sue Hwang;
   Geoscience and Remote Sensing Symposium, 2005. IGARSS '05. Proceedings. 2005 IEEE International
   Volume 2, 25-29 July 2005 Page(s):4 pp.
   IEEE CNF

5. **Interactive ROLAP on large datasets: a case study with UB-trees**
   Ramsak, F.; Markl, V.; Fenk, R.; Bayer, R.; Ruf, T.;
   Database Engineering & Applications, 2001 International Symposium on.
   16-18 July 2001 Page(s):167 - 176
   IEEE CNF

Indexed by Inspec®

Terms used: **reverse star schema** Found **174** of **207,474**

Sort results by: relevance
Display results: expanded form
Save results to a Binder
Search Tips
Open results in a new window
Try an Advanced Search
Try this search in The ACM Guide

Results 1 - 20 of 174 Result page: **1** 2 3 4 5 6 7 8 9 next

Relevance scale

1 Designing data marts for data warehouses

October 2001 **ACM Transactions on Software Engineering and Methodology (TOSEM)**, Volume 10 Issue 4

Publisher: ACM Press

Full text available: pdf(203.43 KB) Additional Information: full citation, abstract, references, citings, index terms, review

Data warehouses are databases devoted to analytical processing. They are used to support decision-making activities in most modern business settings, when complex data sets have to be studied and analyzed. The technology for analytical processing assumes that data are presented in the form of simple data marts, consisting of a well-identified collection of facts and data analysis dimensions (star schema). Despite the wide diffusion of data warehouse technology and concepts, we still miss me ...

Keywords: conceptual modeling, data mart, data warehouse, design method, software quality management

2 Snakes and sandwiches: optimal clustering strategies for a data warehouse

H. V. Jagadish, Laks V. S. Lakshmanan, Divesh Srivastava

June 1999 **ACM SIGMOD Record , Proceedings of the 1999 ACM SIGMOD international conference on Management of data SIGMOD '99**, Volume 28 Issue 2

Publisher: ACM Press

Full text available: pdf(1.47 MB) Additional Information: full citation, abstract, references, citings, index terms

Physical layout of data is a crucial determinant of performance in a data warehouse. The optimal clustering of data on disk, for minimizing expected I/O, depends on the query workload. In practice, we often have a reasonable sense of the likelihood of different classes of queries, e.g., 40% of the queries concern calls made from some specific telephone number in some month. In this paper, we address the problem of finding an optimal clustering of records of ...

3 A comparison of data warehousing methodologies

Arun Sen, Atish P. Sinha

March 2005 **Communications of the ACM**, Volume 48 Issue 3

Publisher: ACM Press

Full text available: pdf(117.81 KB) Additional Information: full citation, abstract, references, index terms

html(28.41 KB)

Using a common set of attributes to determine which methodology to use in a particular data warehousing project.

4 Star graphics: An object-oriented implementation

Daniel E. Lipkie, Steven R. Evans, John K. Newlin, Robert L. Weissman

July 1982 **ACM SIGGRAPH Computer Graphics , Proceedings of the 9th annual conference on Computer graphics and interactive techniques SIGGRAPH '82**, Volume 16 Issue 3

Publisher: ACM Press

Full text available: pdf(955.07 KB) Additional Information: full citation, abstract, references, citings, index terms

The XEROX Star 8010 Information System features an integrated text and graphics editor. The Star hardware consists of a processor, a large bit-mapped display, a keyboard and a pointing device. Star's basic graphic elements are points, lines, rectangles, triangles, graphics frames, text frames and bar charts. The internal representation is in terms of idealized objects that are displayed or printed at resolutions determined by the output device. This paper describes the design and implementa ...

Keywords: Business graphics, Subclassing

5 Heuristic optimization of OLAP queries in multidimensionally hierarchically clustered databases

Dimitri Theodoratos, Aris Tsois

November 2001 **Proceedings of the 4th ACM international workshop on Data warehousing and OLAP DOLAP '01**

Publisher: ACM Press

Full text available: pdf(1.44 MB) Additional Information: full citation, abstract, citings, index terms

On-line analytical processing (OLAP) is a technology that encompasses applications requiring a multidimensional and hierarchical view of data. OLAP applications often require fast response time to complex grouping/aggregation queries on enormous quantities of data. Commercial relational database management systems use mainly multiple one-dimensional indexes to process OLAP queries that restrict multiple dimensions. However, in many cases, multidimensional access methods outperform one-dimensiona ...

6 Bottom-up computation of sparse and Iceberg CUBE

Kevin Beyer, Raghu Ramakrishnan

June 1999 **ACM SIGMOD Record , Proceedings of the 1999 ACM SIGMOD international conference on Management of data SIGMOD '99**, Volume 28 Issue 2

Publisher: ACM Press

Full text available: pdf(1.49 MB) Additional Information: full citation, abstract, references, citings, index terms

We introduce the Iceberg-CUBE problem as a reformulation of the datacube (CUBE) problem. The Iceberg-CUBE problem is to compute only those group-by partitions with an aggregate value (e.g., count) above some minimum support threshold. The result of Iceberg-CUBE can be used (1) to answer group-by queries with a clause such as HAVING COUNT(*) >= X, where X is greater than the threshold, (2) for mining multidimensional association rules, and (3) to complement existing strategies for identif ...

7 Automated data warehousing for rule-based CRM systems

Han-joon Kim, TaeHee Lee, Sang-goo Lee, Jonghun Chun

January 2003 **Proceedings of the 14th Australasian database conference - Volume 17 ADC '03**
Publisher: Australian Computer Society, Inc.
Full text available: pdf(274.28 KB) Additional Information: full citation, abstract, references, index terms

This paper proposes a novel way of automatically developing data warehouse configuration in rule-based CRM systems. Rule-based CRM systems assume that marketing activities are represented as a set of **IF-WHEN** rules. Currently, to provide good quality CRM functionalities, CRM systems seek to combine conventional CRM methodologies with data warehousing technology. A data warehouse can be abstractly seen as a set of materialized views. Selecting views for materialization in a data warehouse i ...

Keywords: CRM, analysis query, data warehouse, materialized view, rules, star-join index

8 Component-driven engineering of database applications
Klaus-Dieter Schewe, Bernhard Thalheim
January 2006 **Proceedings of the 3rd Asia-Pacific conference on Conceptual modelling - Volume 53 APCCM '06**
Publisher: Australian Computer Society, Inc.
Full text available: pdf(188.64 KB) Additional Information: full citation, abstract, references, index terms

Though it is commonly agreed that the design of large database schemata requires group effort, database design from component subschemata has not been investigated thoroughly. In this paper we investigate snowflake-like subschemata of database schemata expressed in the Higher-order Entity-Relationship Model (HERM). These subschemata are almost hierarchical in the sense that they may contain cycles in the schema, but not in the instances. We show that each HERM schema can be decomposed into such ...

9 Poster papers - short papers: A visual interface technique for exploring OLAP data with coordinated dimension hierarchies
Mark Sifer
November 2003 **Proceedings of the twelfth international conference on Information and knowledge management CIKM '03**
Publisher: ACM Press
Full text available: pdf(272.82 KB) Additional Information: full citation, abstract, references, index terms

Multi-dimensional data occurs in many domains while a wide variety of text based and visual interfaces for querying such data exists. But many of these interfaces are not applicable to OLAP, as they do not support use of dimension hierarchies for selection and aggregation. We introduce an interface technique which supports visual querying of OLAP data, that has been implemented in the SGViewer tool. It is based on a data graph rather than a data cube representation of the data. Our interface pre ...

Keywords: OLAP, data exploration, hierarchies, interface

10 Graphical interaction with heterogeneous databases
T. Catarci, G. Santucci, J. Cardiff
May 1997 **The VLDB Journal — The International Journal on Very Large Data Bases**, Volume 6 Issue 2
Publisher: Springer-Verlag New York, Inc.
Full text available: pdf(602.82 KB) Additional Information: full citation, abstract, citings, index terms

During the past few years our research efforts have been inspired by two different needs.

On one hand, the number of non-expert users accessing databases is growing apace. On the other, information systems will no longer be characterized by a single centralized architecture, but rather by several heterogeneous component systems. In order to address such needs we have designed a new query system with both user-oriented and multidatabase features. The system's main components are an adaptive visua ...

11 Charles W. Bachman interview: September 25-26, 2004; Tucson, Arizona

Thomas Haigh
January 2006 **ACM Oral History interviews**
Publisher: ACM Press
Full text available: pdf(761.66 KB) Additional Information: full citation, abstract

Charles W. Bachman reviews his career. Born during 1924 in Kansas, Bachman attended high school in East Lansing, Michigan before joining the Army Anti Aircraft Artillery Corp, with which he spent two years in the Southwest Pacific Theater, during World War II. After his discharge from the military, Bachman earned a B.Sc. in Mechanical Engineering in 1948, followed immediately by an M.Sc. in the same discipline, from the University of Pennsylvania. On graduation, he went to work for Do ...

12 Session 7: GYO reductions, canonical connections, tree and cyclic schemas and tree projections

Nathan Goodman, Oded Shmueli, Y. C. Tay
March 1983 **Proceedings of the 2nd ACM SIGACT-SIGMOD symposium on Principles of database systems PODS '83**
Publisher: ACM Press
Full text available: pdf(1.09 MB) Additional Information: full citation, abstract, references, citings

Database schemas may be partitioned into two sub-classes tree schemas and cyclic schemas. The analysis of tree vs cyclic schemas introduced the concepts of GYO reductions, canonical connections and tree projections. This paper investigates the intricate relationships among these concepts in the context of universal relation databases.

13 The theory of parsing, translation, and compiling

Alfred V. Aho, Jeffrey D. Ullman
January 1972 Book
Publisher: Prentice-Hall, Inc.
Full text available: pdf(98.28 MB) Additional Information: full citation, abstract, references, cited by, index terms

From volume 1 Preface (See Front Matter for full Preface)

This book is intended for a one or two semester course in compiling theory at the senior or graduate level. It is a theoretically oriented treatment of a practical subject. Our motivation for making it so is threefold.

(1) In an area as rapidly changing as Computer Science, sound pedagogy demands that courses emphasize ideas, rather than implementation details. It is our hope that the algorithms and concepts presen ...

14 Special topic section on peer to peer data management: Design issues and challenges for RDF- and schema-based peer-to-peer systems

Wolfgang Nejdl, Wolf Siberski, Michael Sintek
September 2003 **ACM SIGMOD Record**, Volume 32 Issue 3
Publisher: ACM Press

Full text available: pdf(135.94 KB) Additional Information: full citation, abstract, references, citings

Databases have employed a schema-based approach to store and retrieve structured data for decades. For peer-to-peer (P2P) networks, similar approaches are just beginning to emerge. While quite a few database techniques can be re-used in this new context, a P2P data management infrastructure poses additional challenges which have to be solved before schema-based P2P networks become as common as schema-based databases. We will describe some of these challenges and discuss approaches to solve them. ...

15 Semantics and implementation of schema evolution in object-oriented databases

Jay Banerjee, Won Kim, Hyoung-Joo Kim, Henry F. Korth

December 1987 **ACM SIGMOD Record , Proceedings of the 1987 ACM SIGMOD international conference on Management of data SIGMOD '87**, Volume 16 Issue 3

Publisher: ACM Press

Full text available: pdf(1.54 MB) Additional Information: full citation, abstract, references, citings, index terms

Object-oriented programming is well-suited to such data-intensive application domains as CAD/CAM, AI, and OIS (office information systems) with multimedia documents. At MCC we have built a prototype object-oriented database system, called ORION. It adds persistence and sharability to objects created and manipulated in applications implemented in an object-oriented programming environment. One of the important requirements of these applications is schema evolution, that is, the ability to dy ...

16 Data processing in the large: BIwTL: a business information warehouse toolkit and language for warehousing simplification and automation

Bin He, Rui Wang, Ying Chen, Ana Lelescu, James Rhodes

June 2007 **Proceedings of the 2007 ACM SIGMOD international conference on Management of data SIGMOD '07**

Publisher: ACM Press

Full text available: pdf(355.95 KB) Additional Information: full citation, abstract, references, index terms

Rapidly leveraging information analytics technologies to mine the mounting information in structured and unstructured forms, derive business insights and improve decision making is becoming increasingly critical to today's business successes. One of the key enablers of the analytics technologies is an Information Warehouse Management System (IWMS) that processes different types and forms of information, builds, and maintains the information warehouse (IW) effectively. Although traditional mul ...

Keywords: data mining, information warehouse, warehousing language

17 A graphical definition of authorization schema in the DTAC model

Jonathon E. Tidswell, John M. Potter

May 2001 **Proceedings of the sixth ACM symposium on Access control models and technologies SACMAT '01**

Publisher: ACM Press

Full text available: pdf(186.83 KB) Additional Information: full citation, abstract, references, index terms

The specification of constraint languages for access control models has proven to be difficult but remains necessary for safety and for mandatory access control policies: While the authorisation relation $(Subject \times Object \rightarrow \pow Right)$ defines the authorised permissions an authorisation schema defines how the various concepts (such as subjects, users, roles, labels) are combined to form a complete access control model.Using examples drawn from common access contr ...

Keywords: DTAC, access control, computer security, constraints, dynamic, graphs, roles,

schema, type

18 Computation: finite and infinite machines

Marvin L. Minsky
January 1967 Book
Publisher: Prentice-Hall, Inc.
Additional Information: full citation, abstract, references, cited by, index terms

From the Preface (See Front Matter for full Preface)

Man has within a single generation found himself sharing the world with a strange new species: the computers and computer-like machines. Neither history, nor philosophy, nor common sense will tell us how these machines will affect us, for they do not do "work" as did machines of the Industrial Revolution. Instead of dealing with materials or energy, we are told that they handle "control" and "information" and even "intellectua ...

19 HydroJ: object-oriented pattern matching for evolvable distributed systems

Keunwoo Lee, Anthony LaMarca, Craig Chambers
October 2003 **ACM SIGPLAN Notices , Proceedings of the 18th annual ACM SIGPLAN conference on Object-oriented programing, systems, languages, and applications OOPSLA '03**, Volume 38 Issue 11
Publisher: ACM Press

Full text available: pdf(311.06 KB) Additional Information: full citation, abstract, references, citings, index terms

In an evolving software system, components must be able to change independently while remaining compatible with their peers. One obstacle to independent evolution is the *brittle parameter problem*: the ability of two components to communicate can depend on a number of *inessential* details of the types, structure, and/or contents of the values communicated. If these details change, then the components can no longer communicate, even if the *essential* parts of the message remain ...

Keywords: HydroJ, XML, distributed systems, dynamic dispatch, object-oriented programming, pattern matching, semi-structured data, software evolution, ubiquitous computing

20 XML schemas: integration and translation: NeT & CoT: translating relational schemas to XML schemas using semantic constraints

Dongwon Lee, Murali Mani, Frank Chiu, Wesley W. Chu
November 2002 **Proceedings of the eleventh international conference on Information and knowledge management CIKM '02**
Publisher: ACM Press

Full text available: pdf(321.89 KB) Additional Information: full citation, abstract, references, citings, index terms

Two algorithms, called NeT and CoT, to translate relational schemas to XML schemas using various semantic constraints are presented. The XML schema representation we use is a language-independent formalism named XSchema, that is both precise and concise. A given XSchema can be mapped to a schema in any of the existing XML schema language proposals. Our proposed algorithms have the following characteristics: (1) NeT derives a nested structure from a flat relational model by repeatedly applying th ...

Keywords: XML, schema translation, semantic constraints

Results 1 - 20 of 174 Result page: **1** 2 3 4 5 6 7 8 9 next

Searched for **snowflake schema**. Results **1 - 1** of about **1**. Search took **0.02** seconds.

Sort by: Date / Relevance

Data Warehouse Glossary - IT@CSUMB.EDU

... Snowflake structure: **Snowflake** is a star **schema** with normalized dimensions. Source data: The data from the operational or legacy systems that feed the ETL ...

it.csumb.edu/departments/data/glossary.html - 59k - 2005-10-19 - Cached

snowflake schema CSUMB Search

Powered by Google

An overview of data warehousing and OLAP technology

Full text Pdf (102 KB)

Source **ACM SIGMOD Record** archive
Volume 26 , Issue 1 (March 1997) table of contents
Pages: 65 - 74
Year of Publication: 1997
ISSN:0163-5808

Authors Surajit Chaudhuri Microsoft Research, Redmond
Umeshwar Dayal Hewlett-Packard Labs, Palo Alto

Publisher ACM Press New York, NY, USA

↑ ABSTRACT

Data warehousing and on-line analytical processing (OLAP) are essential elements of decision support, which has increasingly become a focus of the database industry. Many commercial products and services are now available, and all of the principal database management system vendors now have offerings in these areas. Decision support places some rather different requirements on database technology compared to traditional on-line transaction processing applications. This paper provides an overview of data warehousing and OLAP technologies, with an emphasis on their new requirements. We describe back end tools for extracting, cleaning and loading data into a data warehouse; multidimensional data models typical of OLAP; front end client tools for querying and data analysis; server extensions for efficient query processing; and tools for metadata management and for managing the warehouse. In addition to surveying the state of the art, this paper also identifies some promising research issues, some of which are related to problems that the database research community has worked on for years, but others are only just beginning to be addressed. This overview is based on a tutorial that the authors presented at the VLDB Conference, 1996.

↑ CITED BY *153*

I. Petrounias , A. Assaid, Temporal web log mining using olap techniques, Proceedings of the international conference on Computational methods in sciences and engineering, p.512-519, September 12-16, 2003, Kastoria, Greece

Yon Dohn Chung , Woo Suk Yang , Myoung Ho Kim, An efficient, robust method for processing of partial top-k/bottom-k queries using the RD-Tree in OLAP, Decision Support Systems, v.43 n.2, p.313-321, March, 2007

Biswadeep Nag , David J. DeWitt, Memory allocation strategies for complex decision support

queries, Proceedings of the seventh international conference on Information and knowledge management, p.116-123, November 02-07, 1998, Bethesda, Maryland, United States

Seokjin Hong , Jinuk Bae , Taewon Lee , Sukho Lee, Histogram-by: A grouping operator for continuous domains, Data & Knowledge Engineering, v.60 n.3, p.451-467, March, 2007

Anoop Singhal , Sushil Jajodia, Data warehousing and data mining techniques for intrusion detection systems, Distributed and Parallel Databases, v.20 n.2, p.149-166, September 2006

Yeow Wei Choong , Dominique Laurent , Patrick Marcel, Computing appropriate representations for multidimensional data, Data & Knowledge Engineering, v.45 n.2, p.181-203, May 2003

Anoop Singhal, Design of a data warehouse system for network/web services, Proceedings of the thirteenth ACM international conference on Information and knowledge management, November 08-13, 2004, Washington, D.C., USA

Surajit Chaudhuri , Umeshwar Dayal, Data warehousing and OLAP for decision support, ACM SIGMOD Record, v.26 n.2, p.507-508, June 1997

Nenad Jukic , Tania Neild, Designing a first-iteration data warehouse for a financial application service provider, Annals of cases on information technology, Idea Group Publishing, Hershey, PA, 2002

Jiawei Han, Data mining tasks and methods: Rule discovery: characteristic rules, Handbook of data mining and knowledge discovery, Oxford University Press, Inc., New York, NY, 2002

Yeow Wei Choong , Dominique Laurent , Patrick Marcel, Computing appropriate representations for multidimensional data, Proceedings of the 4th ACM international workshop on Data warehousing and OLAP, p.16-23, November 09-09, 2001, Atlanta, Georgia, USA

Carlos Ordonez, Vertical and horizontal percentage aggregations, Proceedings of the 2004 ACM SIGMOD international conference on Management of data, June 13-18, 2004, Paris, France

Fernando Berzal , Ignacio Blanco , Juan-Carlos Cubero , Nicolas Marin, Component-based data mining frameworks, Communications of the ACM, v.45 n.12, December 2002

Vivekanand Gopalkrishnan , Qing Li , Kamalakar Karlapalem, Efficient query processing with structural join indexing in an object relational data warehousing environment, Data warehousing and web engineering, IRM Press, Hershey, PA, 2002

Antoaneta Ivanova , Boris Rachev, Multidimensional models: constructing data CUBE, Proceedings of the 5th international conference on Computer systems and technologies, June 17-18, 2004, Rousse, Bulgaria

Surajit Chaudhuri , Umesh Dayal, Data and knowledge in database systems: multidimensional databases and online analytical processing, Handbook of data mining and knowledge discovery, Oxford University Press, Inc., New York, NY, 2002

Mark Levene , George Loizou, Why is the snowflake schema a good data warehouse design?, Information Systems, v.28 n.3, p.225-240, May 2003

Eugene Inseok Chong , Jagannathan Srinivasan , Souripriya Das , Chuck Freiwald , Aravind Yalamanchi , Mahesh Jagannath , Anh-Tuan Tran , Ramkumar Krishnan , Richard Jiang, A mapping mechanism to support bitmap index and other auxiliary structures on tables stored as primary B^{+}-trees, ACM SIGMOD Record, v.32 n.2, June 2003

Lingli Ding , Xin Zhang , Elke A. Rundensteiner, The MRE wrapper approach: enabling incremental view maintenance of data warehouses defined on multi-relation information sources, Proceedings of the 2nd ACM international workshop on Data warehousing and OLAP, p.30-35, November 02-06, 1999, Kansas City, Missouri, United States

- Antonio Badia , Matt Chanda , Bin Cao, Adding subqueries to MySQL, what does it take to have a decision-support engine?, Proceedings of the 5th ACM international workshop on Data Warehousing and OLAP, p.49-56, November 08-08, 2002, McLean, Virginia, USA
- David W. Cheung , Bo Zhou , Ben Kao , Hongjun Lu , Tak Wah Lam , Hing Fung Ting, Requirement-based data cube schema design, Proceedings of the eighth international conference on Information and knowledge management, p.162-169, November 02-06, 1999, Kansas City, Missouri, United States
- Eugene Inseok Chong , Jagannathan Srinivasan , Souripriya Das , Chuck Freiwald , Aravind Yalamanchi , Mahesh Jagannath , Anh-Tuan Tran , Ramkumar Krishnan , Richard Jiang, A mapping mechanism to support bitmap index and other auxiliary structures on tables stored as primary $B^{\pm}$-trees, Proceedings of the eleventh international conference on Information and knowledge management, November 04-09, 2002, McLean, Virginia, USA
- Sriram Padmanabhan , Bishwaranjan Bhattacharjee , Tim Malkemus , Leslie Cranston , Matthew Huras, Multi-dimensional clustering: a new data layout scheme in DB2, Proceedings of the 2003 ACM SIGMOD international conference on Management of data, June 09-12, 2003, San Diego, California

  Andrew N. K. Chen , Paulo B. Goes , Alok Gupta , James R. Marsden, Database design in the modern organization: identifying robust structures under changing query patterns and arrival rate conditions, Decision Support Systems, v.37 n.3, p.435-447, June 2004
- Satyadeep Patnaik , Marshall Meier , Brian Henderson , Joe Hickman , Brajendra Panda, Improving the performance of lineage tracing in data warehouse, Proceedings of the 1999 ACM symposium on Applied computing, p.210-215, February 28-March 02, 1999, San Antonio, Texas, United States
- Riadh Ben Messaoud , Omar Boussaid , Sabine Rabaséda, A new OLAP aggregation based on the AHC technique, Proceedings of the 7th ACM international workshop on Data warehousing and OLAP, November 12-13, 2004, Washington, DC, USA
- Riadh Ben Messaoud , Omar Boussaid , Sabine Loudcher Rabaséda, Evaluation of a MCA-based approach to organize data cubes, Proceedings of the 14th ACM international conference on Information and knowledge management, October 31-November 05, 2005, Bremen, Germany

  Jiawei Han , Laks V. S. Lakshmanan , Raymond T. Ng, Constraint-Based, Multidimensional Data Mining, Computer, v.32 n.8, p.46-50, August 1999

  Ming-Chuan Hung , Man-Lin Huang , Don-Lin Yang , Nien-Lin Hsueh, Efficient approaches for materialized views selection in a data warehouse, Information Sciences: an International Journal, v.177 n.6, p.1333-1348, March, 2007

  Nebojsa Stefanovic , Jiawei Han , Krzysztof Koperski, Object-Based Selective Materialization for Efficient Implementation of Spatial Data Cubes, IEEE Transactions on Knowledge and Data Engineering, v.12 n.6, p.938-958, November 2000
- Dimitri Theodoratos, Exploiting hierarchical clustering in evaluating multidimensional aggregation queries, Proceedings of the 6th ACM international workshop on Data warehousing and OLAP, November 07-07, 2003, New Orleans, Louisiana, USA

  Nikos Karayannidis , Timos Sellis, SISYPHUS: the implementation of a chunk-based storage manager for OLAP data cubes, Data & Knowledge Engineering, v.45 n.2, p.155-180, May 2003

  Edward Hung , David W. Cheung , Ben Kao, Optimization in Data Cube System Design, Journal of Intelligent Information Systems, v.23 n.1, p.17-45, July 2004
- John Horner , Il-Yeol Song , Peter P. Chen, An analysis of additivity in OLAP systems, Proceedings of the 7th ACM international workshop on Data warehousing and OLAP, November

12-13, 2004, Washington, DC, USA

Nam Huyn, Data analysis and mining in the life sciences, ACM SIGMOD Record, v.30 n.3, September 2001

Chang-Sup Park , Myoung Ho Kim , Yoon-Joon Lee, Usability-based caching of query results in OLAP systems, Journal of Systems and Software, v.68 n.2, p.103-119, 15 November 2003

Chee-Yong Chan , Yannis E. Ioannidis, Bitmap index design and evaluation, ACM SIGMOD Record, v.27 n.2, p.355-366, June 1998

Osmar R. Zaïane , Jiawei Han , Ze-Nian Li , Jean Hou, Mining multimedia data, Proceedings of the 1998 conference of the Centre for Advanced Studies on Collaborative research, p.24, November 30-December 03, 1998, Toronto, Ontario, Canada

Jiawei Han , Jenny Y. Chiang , Sonny Chee , Jianping Chen , Qing Chen , Shan Cheng , Wan Gong , Micheline Kamber , Krzysztof Koperski , Gang Liu , Yijun Lu , Nebojsa Stefanovic , Lara Winstone , Betty B. Xia , Osmar R. Zaiane , Shuhua Zhang , Hua Zhu, DBMiner: a system for data mining in relational databases and data warehouses, Proceedings of the 1997 conference of the Centre for Advanced Studies on Collaborative research, p.8, November 10-13, 1997, Toronto, Ontario, Canada

Mirek Riedewald , Divyakant Agrawal , Amr El Abbadi, Dynamic multidimensional data cubes, Multidimensional databases: problems and solutions, Idea Group Publishing, Hershey, PA, 2003

L. Palopoli , L. Pontieri , G. Terracina , D. Ursino, A novel three-level architecture for large data warehouses, Journal of Systems Architecture: the EUROMICRO Journal, v.47 n.11, p.937-958, May 2002

Yingwei Cui , Jennifer Widom , Janet L. Wiener, Tracing the lineage of view data in a warehousing environment, ACM Transactions on Database Systems (TODS), v.25 n.2, p.179-227, June 2000

Amin Y. Noaman , Ken Barker, A horizontal fragmentation algorithm for the fact relation in a distributed data warehouse, Proceedings of the eighth international conference on Information and knowledge management, p.154-161, November 02-06, 1999, Kansas City, Missouri, United States

Anindya Datta , Debra VanderMeer , Krithi Ramamritham, Parallel Star Join + DataIndexes: Efficient Query Processing in Data Warehouses and OLAP, IEEE Transactions on Knowledge and Data Engineering, v.14 n.6, p.1299-1316, November 2002

Chung-Min Chen , Munir Cochinwala , Elsa Yueh, Dealing with slow-evolving fact: a case study on inventory data warehousing, Proceedings of the 2nd ACM international workshop on Data warehousing and OLAP, p.22-29, November 02-06, 1999, Kansas City, Missouri, United States

Roberto Esposito , Rosa Meo , Marco Botta, Answering constraint-based mining queries on itemsets using previous materialized results, Journal of Intelligent Information Systems, v.26 n.1, p.95-111, January 2006

Sam Y. Sung , Zhao Li , Peng Sun, A fast filtering scheme for large database cleansing, Proceedings of the eleventh international conference on Information and knowledge management, November 04-09, 2002, McLean, Virginia, USA

Pauline LienHua Chou , Xiuzhen Zhang, Efficiently computing the top *N* averages in iceberg cubes, Proceedings of the twenty-sixth Australasian conference on Computer science: research and practice in information technology, p.101-109, February 01, 2003, Adelaide, Australia

Nenad Jukic, Modeling strategies and alternatives for data warehousing projects, Communications of the ACM, v.49 n.4, p.83-88, April 2006

Sham Prasher , Xiaofang Zhou, Multiresolution amalgamation: dynamic spatial data cube generation, Proceedings of the fifteenth Australasian database conference, p.103-111, January 01, 2004, Dunedin, New Zealand

José María Cavero , Esperanza Marcos , Mario Piattini , Adolfo Sánchez, A methodology for datawarehouse design: conceptual modeling, Data warehousing and web engineering, IRM Press, Hershey, PA, 2002

Cecil Eng Huang Chua , Roger H. L. Chiang , Ee-Peng Lim, An intelligent middleware for linear correlation discovery, Decision Support Systems, v.32 n.4, p.313-326, March 2002

Chang-Sup Park , Myoung Ho Kim , Yoon-Joon Lee, Finding an efficient rewriting of OLAP queries using materialized views in data warehouses, Decision Support Systems, v.32 n.4, p.379-399, March 2002

Swarup Acharya , Phillip B. Gibbons , Viswanath Poosala, Congressional samples for approximate answering of group-by queries, ACM SIGMOD Record, v.29 n.2, p.487-498, June 2000

Young-Koo Lee , Kyu-Young Whang , Yang-Sae Moon , Il-Yeol Song, An aggregation algorithm using a multidimensional file in multidimensional OLAP, Information Sciences: an International Journal, v.152 n.1, p.121-138, June 2003

Venkatesh Ganti , Johannes Gehrke , Raghu Ramakrishnan, DEMON: Mining and Monitoring Evolving Data, IEEE Transactions on Knowledge and Data Engineering, v.13 n.1, p.50-63, January 2001

Avigdor Gal , Vijayalakshmi Atluri, An authorization model for temporal data, Proceedings of the 7th ACM conference on Computer and communications security, p.144-153, November 01-04, 2000, Athens, Greece

Amit Rudra , Raj Gopalan, Adaptive use of a cluster of PCs for data warehousing applications: some problems and issues, Proceedings of the 2000 ACM symposium on Applied computing, p.698-703, March 2000, Como, Italy

Sara Cohen, Containment of aggregate queries, ACM SIGMOD Record, v.34 n.1, March 2005

L. Schlesinger , A. Bauer , W. Lehner , G. Ediberidze , M. Gutzmann, Efficiently synchronizing multidimensional schema data, Proceedings of the 4th ACM international workshop on Data warehousing and OLAP, p.69-76, November 09-09, 2001, Atlanta, Georgia, USA

Sergio Greco , Luigi Pontieri , Ester Zumpano, A technique for information system integration, Proceedings of the 2001 international conference on Information systems technology and its applications, p.75-84, June 13-15, 2001, Kharkiv, Ukraine

Jiawei Han , Jian Pei , Guozhu Dong , Ke Wang, Efficient computation of Iceberg cubes with complex measures, ACM SIGMOD Record, v.30 n.2, p.1-12, June 2001

Erik Riedel , Christos Faloutsos , Gregory R. Ganger , David F. Nagle, Data mining on an OLTP system (nearly) for free, ACM SIGMOD Record, v.29 n.2, p.13-21, June 2000

J. Trujillo , M. Palomar, An object oriented approach to multidimensional database conceptual modeling (OOMD), Proceedings of the 1st ACM international workshop on Data warehousing and OLAP, p.16-21, November 02-07, 1998, Washington, D.C., United States

Dimitri Theodoratos , Timos Sellis, Incremental Design of a Data Warehouse, Journal of Intelligent Information Systems, v.15 n.1, p.7-27, July—Aug. 2000

Rami Rifaieh , Nabila Aïcha Benharkat, Query-based data warehousing tool, Proceedings of the 5th ACM international workshop on Data Warehousing and OLAP, p.35-42, November 08-08,

2002, McLean, Virginia, USA

Torsten Priebe , Günther Pernul, Towards integrative enterprise knowledge portals, Proceedings of the twelfth international conference on Information and knowledge management, November 03-08, 2003, New Orleans, LA, USA

Bongki Moon , Ines Fernando Vega Lopez , Vijaykumar Immanuel, Efficient Algorithms for Large-Scale Temporal Aggregation, IEEE Transactions on Knowledge and Data Engineering, v.15 n.3, p.744-759, March 2003

Roland Holten, Specification of management views in information warehouse projects, Information Systems, v.28 n.7, p.709-751, 1 October 2003

Luigi Palopoli , Giorgio Terracina , Domenico Ursino, Experiences using DIKE, a system for supporting cooperative information system and data warehouse design, Information Systems, v.28 n.7, p.835-865, 1 October 2003

Hao Fan , Alexandra Poulovassilis, Using AutoMed metadata in data warehousing environments, Proceedings of the 6th ACM international workshop on Data warehousing and OLAP, November 07-07, 2003, New Orleans, Louisiana, USA

Mirek Riedewald , Divyakant Agrawal , Amr El Abbadi, Efficient integration and aggregation of historical information, Proceedings of the 2002 ACM SIGMOD international conference on Management of data, June 03-06, 2002, Madison, Wisconsin

Yannis Sismanis , Antonios Deligiannakis , Yannis Kotidis , Nick Roussopoulos, Hierarchical dwarfs for the rollup cube, Proceedings of the 6th ACM international workshop on Data warehousing and OLAP, November 07-07, 2003, New Orleans, Louisiana, USA

Yannis Kotidis , Nick Roussopoulos, A case for dynamic view management, ACM Transactions on Database Systems (TODS), v.26 n.4, p.388-423, December 2001

Chien-Le Goh , Kazuki Aisaka , Masahiko Tsukamoto , Shojiro Nishio, Database compression with data mining methods, Information organization and databases: foundations of data organization, Kluwer Academic Publishers, Norwell, MA, 2000

Rajesh R. Bordawekar , Christian A. Lang, Analytical processing of XML documents: opportunities and challenges, ACM SIGMOD Record, v.34 n.2, June 2005

Ladjel Bellatreche , Arnaud Giacometti , Patrick Marcel , Hassina Mouloudi , Dominique Laurent, A personalization framework for OLAP queries, Proceedings of the 8th ACM international workshop on Data warehousing and OLAP, November 04-05, 2005, Bremen, Germany

Doron Rotem , Kurt Stockinger , Kesheng Wu, Optimizing candidate check costs for bitmap indices, Proceedings of the 14th ACM international conference on Information and knowledge management, October 31-November 05, 2005, Bremen, Germany

Gang Gou , Maxim Kormilitsin , Rada Chirkova, Query evaluation using overlapping views: completeness and efficiency, Proceedings of the 2006 ACM SIGMOD international conference on Management of data, June 27-29, 2006, Chicago, IL, USA

Rada Chirkova , Fereidoon Sadri, Query optimization using restructured views, Proceedings of the 15th ACM international conference on Information and knowledge management, November 06-11, 2006, Arlington, Virginia, USA

Anne-Muriel Arigon , Anne Tchounikine , Maryvonne Miquel, A multiversion model for multimedia data warehouse, Proceedings of the 6th international workshop on Multimedia data mining: mining integrated media and complex data, p.7-13, August 21-21, 2005, Chicago, Illinois

- Marcelo Arenas , Leopoldo Bertossi , Jan Chomicki, Consistent query answers in inconsistent databases, Proceedings of the eighteenth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems, p.68-79, May 31-June 03, 1999, Philadelphia, Pennsylvania, United States
- Franck Ravat , Olivier Teste , Giles Zurfluh, Towards data warehouse design, Proceedings of the eighth international conference on Information and knowledge management, p.359-366, November 02-06, 1999, Kansas City, Missouri, United States
- Rónán Páircéir , Sally McClean , Bryan Scotney, Discovery of multi-level rules and exceptions from a distributed database, Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining, p.523-532, August 20-23, 2000, Boston, Massachusetts, United States
- Nguyen Thanh Binh , A. Min Tjoa, Conceptual multidimensional data model based on object-oriented metacube, Proceedings of the 2001 ACM symposium on Applied computing, p.295-300, March 2001, Las Vegas, Nevada, United States

Mario Cannataro , Domenico Talia , Pradip K. Srimani, Parallel data intensive computing in scientific and commercial applications, Parallel Computing, v.28 n.5, p.673-704, May 2002

Jorge R. Bernardino , Pedro S. Furtado , Henrique C. Madeira, Approximate Query Answering Using Data Warehouse Striping, Journal of Intelligent Information Systems, v.19 n.2, p.145-167, September 2002

Timo Niemi , Lasse Hirvonen , Kalervo Järvelin, Multidimensional data model and query language for informetrics, Journal of the American Society for Information Science and Technology, v.54 n.10, p.939-951, August 2003

L. Palopoli , D. Saccà , G. Terracina , D. Ursino, A technique for deriving hyponymies and overlappings from database schemes, Data & Knowledge Engineering, v.40 n.3, p.285-314, March 2002

Joachim Hammer , Lixin Fu, CubiST^{++}: Evaluating Ad-Hoc CUBE Queries Using Statistics Trees, Distributed and Parallel Databases, v.14 n.3, p.221-254, November 2003

Reda Alhajj , Mehmet Kaya, Employing OLAP mining for multiagent reinforcement learning, Design and application of hybrid intelligent systems, IOS Press, Amsterdam, The Netherlands, 2003

Guozhu Dong , Jiawei Han , Joyce M. W. Lam , Jian Pei , Ke Wang , Wei Zou, Mining Constrained Gradients in Large Databases, IEEE Transactions on Knowledge and Data Engineering, v.16 n.8, p.922-938, August 2004

Hai Wang , Kenneth C. Sevcik, A multi-dimensional histogram for selectivity estimation and fast approximate query answering, Proceedings of the 2003 conference of the Centre for Advanced Studies on Collaborative research, p.328-342, October 06-09, 2003, Toronto, Ontario, Canada

Jiawei Han , Yongjian Fu, Mining Multiple-Level Association Rules in Large Databases, IEEE Transactions on Knowledge and Data Engineering, v.11 n.5, p.798-805, September 1999

- Kristen LeFevre , David J. DeWitt , Raghu Ramakrishnan, Incognito: efficient full-domain K-anonymity, Proceedings of the 2005 ACM SIGMOD international conference on Management of data, June 14-16, 2005, Baltimore, Maryland
- Tho Manh Nguyen , Josef Schiefer , A. Min Tjoa, Sense & response service architecture (SARESA): an approach towards a real-time business intelligence solution and its use for a fraud detection application, Proceedings of the 8th ACM international workshop on Data warehousing and OLAP, November 04-05, 2005, Bremen, Germany

David W. Cheung , H. Y. Hwang , Ada W. Fu , Jiawei Han, Efficient Rule-Based Attribute-Oriented Induction for Data Mining, Journal of Intelligent Information Systems, v.15 n.2, p.175-200, Sept./Oct. 2000

Bartosz B□bel , Johann Eder , Christian Koncilia , Tadeusz Morzy , Robert Wrembel, Creation and management of versions in multiversion data warehouse, Proceedings of the 2004 ACM symposium on Applied computing, March 14-17, 2004, Nicosia, Cyprus

Kesheng Wu , Ekow J. Otoo , Arie Shoshani, Optimizing bitmap indices with efficient compression, ACM Transactions on Database Systems (TODS), v.31 n.1, p.1-38, March 2006

Rada Chirkova , Fereidoon Sadri, Query optimization using restructured views, Proceedings of the 15th ACM international conference on Information and knowledge management, November 06-11, 2006, Arlington, Virginia, USA

Ahmed E. Hassan , Richard C. Holt, Architecture recovery of web applications, Proceedings of the 24th International Conference on Software Engineering, May 19-25, 2002, Orlando, Florida

Lixin Fu , Joachim Hammer, CubiST: a new algorithm for improving the performance of ad-hoc OLAP queries, Proceedings of the 3rd ACM international workshop on Data warehousing and OLAP, p.72-79, November 06-11, 2000, McLean, Virginia, United States

Yannis Kotidis, Aggregate view management in data warehouses, Handbook of massive data sets, Kluwer Academic Publishers, Norwell, MA, 2002

Alin Dobra , Minos Garofalakis , Johannes Gehrke , Rajeev Rastogi, Processing complex aggregate queries over data streams, Proceedings of the 2002 ACM SIGMOD international conference on Management of data, June 03-06, 2002, Madison, Wisconsin

Kurt Stockinger , Kesheng Wu , Arie Shoshani, Strategies for processing ad hoc queries on large data warehouses, Proceedings of the 5th ACM international workshop on Data Warehousing and OLAP, p.72-79, November 08-08, 2002, McLean, Virginia, USA

Marco Costa , Henrique Madeira, Handling big dimensions in distributed data warehouses using the DWS technique, Proceedings of the 7th ACM international workshop on Data warehousing and OLAP, November 12-13, 2004, Washington, DC, USA

Surajit Chaudhuri, An overview of query optimization in relational systems, Proceedings of the seventeenth ACM SIGACT-SIGMOD-SIGART symposium on Principles of database systems, p.34-43, June 01-04, 1998, Seattle, Washington, United States

Nenad Jukić , Svetlozar Nestorov, Comprehensive data warehouse exploration with qualified association-rule mining, Decision Support Systems, v.42 n.2, p.859-878, November 2006

Sergio Luján-Mora , Juan Trujillo , Il-Yeol Song, A UML profile for multidimensional modeling in data warehouses, Data & Knowledge Engineering, v.59 n.3, p.725-769, December 2006

Robert M. Bruckner , A. Min Tjoa, Capturing Delays and Valid Times in Data Warehouses—Towards Timely Consistent Analyses, Journal of Intelligent Information Systems, v.19 n.2, p.169-190, September 2002

Esther Pacitti , Pascale Minet , Eric Simon, Replica Consistency in Lazy Master Replicated Databases, Distributed and Parallel Databases, v.9 n.3, p.237-267, May 2001

Sohail Asghar , Damminda Alahakoon, A hybrid neural network based DBMS system for enhanced functionality, Design and application of hybrid intelligent systems, IOS Press, Amsterdam, The Netherlands, 2003

Xinjian Lu , Franklin Lowenthal, Arranging fact table records in a data warehouse to improve query performance, Computers and Operations Research, v.31 n.13, p.2165-2182, November 2004

Sohail Asghar , Damminda Alahakoon , Arthur Hsu, Enhancing OLAP functionality using self-organizing neural networks, Neural, Parallel & Scientific Computations, v.12 n.1, p.1-20, March 2004

Zhiyong Peng , Qing Li , Ling Feng , Xuhui Li , Junqiang Liu, Using Object Deputy Model to Prepare Data for Data Warehousing, IEEE Transactions on Knowledge and Data Engineering, v.17 n.9, p.1274-1288, September 2005

Adam Fadlalla, An experimental investigation of the impact of aggregation on the performance of data mining with logistic regression, Information and Management, v.42 n.5, p.695-707, July 2005

Anne-Muriel Arigon , Anne Tchounikine , Maryvonne Miquel, Handling multiple points of view in a multimedia data warehouse, ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP), v.2 n.3, p.199-218, August 06

Zina Ben-Miled , Yang Liu , Michael Bem , Robert Jones , Robert Oppelt , Samuel Milosevich , Dave Powers , Omran Bukhres, Data access performance in a large and dynamic pharmaceutical drug candidate database, Proceedings of the 2000 ACM/IEEE conference on Supercomputing (CDROM), p.22-es, November 04-10, 2000, Dallas, Texas, United States

Josiane Mothe , Claude Chrisment , Bernard Dousset , Joel Alaux, DocCube: multi-dimensional visualisation and exploration of large document sets, Journal of the American Society for Information Science and Technology, v.54 n.7, p.650-659, May 2003

Designing data marts for data warehouses, ACM Transactions on Software Engineering and Methodology (TOSEM), v.10 n.4, p.452-483, Oct. 2001

Ling Feng , Tharam S. Dillon, Using Fuzzy Linguistic Representations to Provide Explanatory Semantics for Data Warehouses, IEEE Transactions on Knowledge and Data Engineering, v.15 n.1, p.86-102, January 2003

Jiawei Han , Kevin Chang, Data Mining for Web Intelligence, Computer, v.35 n.11, p.64-70, November 2002

Song Lin , Donald E. Brown, An outlier-based data association method for linking criminal incidents, Decision Support Systems, v.41 n.3, p.604-615, March 2006

Jiawei Han , Yixin Chen , Guozhu Dong , Jian Pei , Benjamin W. Wah , Jianyong Wang , Y. Dora Cai, Stream Cube: An Architecture for Multi-Dimensional Analysis of Data Streams, Distributed and Parallel Databases, v.18 n.2, p.173-197, September 2005

Stéphane Grumbach , Leonardo Tininini, On the content of materialized aggregate views, Journal of Computer and System Sciences, v.66 n.1, p.133-168, 01 February 2003

Ricardo Chalmeta , Reyes Grangel, ARDIN extension for virtual enterprise integration, Journal of Systems and Software, v.67 n.3, p.141-152, 15 September 2003

Scott Nicholson, The basis for bibliomining: frameworks for bringing together usage-based data mining and bibliometrics through data warehousing in digital library services, Information Processing and Management: an International Journal, v.42 n.3, p.785-804, May 2006

Nicolas Prat , Jacky Akoka , Isabelle Comyn-Wattiau, A UML-based data warehouse design method, Decision Support Systems, v.42 n.3, p.1449-1473, December 2006

Riccardo Torlone, Conceptual multidimensional models, Multidimensional databases: problems and solutions, Idea Group Publishing, Hershey, PA, 2003

Egemen Tanin , Ben Shneiderman , Hairuo Xie, Browsing large online data tables using generalized query previews, Information Systems, v.32 n.3, p.402-423, May, 2007

Yannis Kotidis, Extending the data warehouse for service provisioning data, Data & Knowledge Engineering, v.59 n.3, p.700-724, December 2006

Satyanarayana R Valluri , Soujanya Vadapalli , Kamalakar Karlapalem, View relevance driven materialized view selection in data warehousing environment, Australian Computer Science Communications, v.24 n.2, p.187-196, January-February 2002

Marcelo Arenas , Leopoldo Bertossi, Hypothetical Temporal Reasoning in Databases, Journal of Intelligent Information Systems, v.19 n.2, p.231-259, September 2002


Surajit Chaudhuri , Vivek Narasayya , Sunita Sarawagi, Extracting predicates from mining models for efficient query evaluation, ACM Transactions on Database Systems (TODS), v.29 n.3, p.508-544, September 2004

Dong Xin , Jiawei Han , Hong Cheng , Xiaolei Li, Answering top-k queries with multi-dimensional selections: the ranking cube approach, Proceedings of the 32nd international conference on Very large data bases, September 12-15, 2006, Seoul, Korea

Nick Bassiliades , Ioannis Vlahavas , Ahmed K. Elmagarmid , Elias N. Houstis, InterBase-KB: Integrating a Knowledge Base System with a Multidatabase System for Data Warehousing, IEEE Transactions on Knowledge and Data Engineering, v.15 n.5, p.1188-1205, September 2003

Mirek Riedewald , Divyakant Agrawal , Amr El Abbadi, Managing and analyzing massive data sets with data cubes, Handbook of massive data sets, Kluwer Academic Publishers, Norwell, MA, 2002

Michael O. Akinde , Michael H. Böhlen , Theodore Johnson , Laks V. S. Lakshmanan , Divesh Srivastava, Efficient OLAP query processing in distributed data warehouses, Information Systems, v.28 n.1-2, p.111-135, 01 March 2003


Gediminas Adomavicius , Ramesh Sankaranarayanan , Shahana Sen , Alexander Tuzhilin, Incorporating contextual information in recommender systems using a multidimensional approach, ACM Transactions on Information Systems (TOIS), v.23 n.1, p.103-145, January 2005


Renata Teixeira , Aman Shaikh , Tim Griffin , Geoffrey M. Voelker, Network sensitivity to hot-potato disruptions, ACM SIGCOMM Computer Communication Review, v.34 n.4, October 2004


Juan Trujillo , Manuel Palomar , Jaime Gómez, The GOLD definition language (GDL): an object oriented formal specification language for multidimensional databases, Proceedings of the 2000 ACM symposium on Applied computing, p.346-350, March 2000, Como, Italy

Thomas Thalhammer , Michael Schrefl, Realizing active data warehouses with off-the-shelf database technology, Software—Practice & Experience, v.32 n.12, p.1193-1222, October 2002

Luigi Palopoli , Domenico Rosaci , Giorgio Terracina , Domenico Ursino, A graph-based approach for extracting terminological properties from information sources with heterogeneous formats, Knowledge and Information Systems, v.8 n.4, p.462-497, November 2005


Juan Trujillo , Manuel Palomar , Jaime Gómez, Detecting patterns and OLAP operations in the GOLD model, Proceedings of the 2nd ACM international workshop on Data warehousing and OLAP, p.48-53, November 02-06, 1999, Kansas City, Missouri, United States

Theodore Johnson, Data warehousing, Handbook of massive data sets, Kluwer Academic Publishers, Norwell, MA, 2002

Gediminas Adomavicius , Alexander Tuzhilin, Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions, IEEE Transactions on

Knowledge and Data Engineering, v.17 n.6, p.734-749, June 2005

Leonardo Tininini, Querying multidimensional data, Multidimensional databases: problems and solutions, Idea Group Publishing, Hershey, PA, 2003

Oliver Günther, Data management in environmental information systems, Handbook of massive data sets, Kluwer Academic Publishers, Norwell, MA, 2002

Brenton Louie , Peter Mork , Fernando Martin-Sanchez , Alon Halevy , Peter Tarczy-Hornoch, Methodological Review: Data integration and genomic medicine, Journal of Biomedical Informatics, v.40 n.1, p.5-16, February, 2007

Hanan Samet, Object-based and image-based object representations, ACM Computing Surveys (CSUR), v.36 n.2, p.159-217, June 2004

Klaus R. Dittrich , Hans Fritschi , Stella Gatziu , Andreas Geppert , Anca Vaduva, SAMOS in hindsight: experiences in building an active object-oriented DBMS, Information Systems, v.28 n.5, p.369-392, July 2003

Zhiyuan Chen , Chen Li , Jian Pei , Yufei Tao , Haixun Wang , Wei Wang , Jiong Yang , Jun Yang , Donghui Zhang, Recent progress on selected topics in database research: a report by nine young Chinese researchers working in the United States, Journal of Computer Science and Technology, v.18 n.5, p.538-552, September 2003

↑ INDEX TERMS

Primary Classification:
H. Information Systems
↳ **H.4** INFORMATION SYSTEMS APPLICATIONS
↳ **H.4.2** Types of Systems
↳ **Subjects:** Decision support (e.g., MIS)

Additional Classification:
A. General Literature

General Terms:
Design, Management, Performance, Theory

↑ Collaborative Colleagues:

Surajit Chaudhuri:	Ashraf Aboulnaga	Usama M. Fayyad	Paul Larson	Sunita Sarawagi
	Rakesh Agrawal	Prasanna Ganesan	Andrew Layman	Cyrus Shahabi
	Sanjay Agrawal	Kris Ganjam	Guy M. Lohman	Kyuseok Shim
	Brian Babcock	Venkatesh Ganti	David Lomet	Kyuseok Shim
	Roger Barga	Venky Ganti	David B. Lomet	Abraham Silberschatz
	Jeff Bernhardt	Hector Garcia-Molina	David Madigan	Utkarsh Srivastava
	Adam Bosworth	Shahram Ghandeharizadeh	David Maier	Michael Stonebraker
	Nicolas Bruno	Luis Gravano	Arun Marathe	Manoj Syamala
	Kaushik Chakrabarti	Jim Gray	Amelie Marian	Dilys Thomas
	Ashok K. Chandra	Ashish Kumar Gupta	Rajeev Motwani	Moshe Y. Vardi
	Moses Charikar	Waqar Hasan	Vivek Narasayya	Theo Vassilakis
	Zhiyuan Chen	Vagelis Hristidis	Vivek R. Narasayya	Murali Venkatrao
	Zhiyuan Chen	Seung-won Hwang	Jeffrey F. Naughton	Michael G. Walker
		Arnd Christian König	Amir Netz	Gerhard Weikum
		Rahul Kapoor	Frank Pellow	
			Hamid Pirahesh	

	Nilesh Dalvi Gautam Das Mayur Datar Umesh Dayal Umeshwar Dayal Klaus R. Dittrich Usama Fayyad	Raghav Kaushik Phokion G. Kolaitis Lubor Kollar Henry F. Korth Ravi Krishnamurthy	Spyros Potamianos Ravishankar Ramamurthy Don Reichart Andreas Reuter	Gio Wiederhold Marianne Winslett Yuqing Wu Yuqing Wu Tak W. Yan
Umeshwar Dayal:	M. Tamer Özsu Tamer Özsu Serge Abiteboul Rakesh Agrawal Philip A Bernstein Philip A. Bernstein Joachim Biskup Angela Bonifati Alejandro P. Buchmann Fabio Casati Malu Castellanos Stefano Ceri Arvola Chan Ashok K. Chandra Surajit Chaudhuri Qiming Chen Parvathi Chundi James Clifford Daniel Cotting Kemal A. Delic Klaus Dittrich Klaus R. Dittrich Laurent Douillet Weimin Du Ramez Elmasri	Stephen Fox Venkatesh Ganti Hector Garcia-Molina Pankaj Garg Nathan Goodman Mohamed G. Gouda Jim Gray Daniela Grigori Martin Griss Martin L. Griss Markus Gross Jiawei Han Eric Hanson Ming C. Hao Sandra Heiler Thomas Holenstein M. Hsu Mei-Chun Hsu Meichun Hsu Hai Yann Hwang Hai-Yann Hwang Matthew Jacobsen Magdi N. Kamel Randy H. Katz Daniel A. Keim	Johannes Klein Ravindran Krishnamurthy Adrian Krug Rivka Ladin Terry Landers Vijay Machiraju Frank Manola Dennis McCarthy Dennis R. McCarthy Behzad Mortazavi-Asl Inderpal S. Mumick Erich J. Neuhold M. Tamer Ozsu Jian Pei Barbara Pernici Helen Pinto Niki Pissinou Andreas Reuter Daniel Ries Daniel R. Ries Arnon Rosenthal Mehmet Sayal Jörn Schneidewind Joern Schneidewind Arie Segev	Eric Shan Ming-Chien Shan Dale Skeen John M Smith John Miles Smith Richard Thomas Snodgrass Michael Stonebraker Babis Theodoulidis Douglas Tolbert Shalom Tsur Patrick Valduriez Moshe Y. Vardi Gilbert Vidal Jianyong Wang Ke Wang Gerhard Weikum Jennifer Widom Gio Wiederhold Peter Wright Gene T. J. Wuu Tak W Yan Tak W. Yan Tak Woon Yan Bin Zhang

↑ **Peer to Peer - Readers of this Article have also read:**

- Constructing reality **Proceedings of the 11th annual international conference on Systems documentation**
  Douglas A. Powell , Norman R. Ball , Mansel W. Griffiths

- Data structures for quadtree approximation and compression **Communications of the ACM** 28, 9
  Hanan Samet

- A hierarchical single-key-lock access control using the Chinese remainder theorem **Proceedings of the 1992 ACM/SIGAPP Symposium on Applied computing**
  Kim S. Lee , Huizhu Lu , D. D. Fisher

- The GemStone object database management system **Communications of the ACM** 34, 10
  Paul Butterworth , Allen Otis , Jacob Stein

- An intelligent component database for behavioral synthesis **Proceedings of the 27th**

ACM/IEEE conference on Design automation
Gwo-Dong Chen , Daniel D. Gajski

Data Warehouse Glossary

This glossary is a compilation of definitions contributed by the experts.

Access: The act of retrieving data from the data warehouse databases.

Access Path: The path selected by the database management system to locate and retrieve requested data.

Ad hoc query: A request for information that is normally fabricated and run a single time and cannot be anticipated in advance. It consists of an SQL statement that has been constructed by a knowledgeable user or through a data access tool.

Aggregation: The process by which data values are collected with the intent to manage the collection as a single unit.
Example: The combination of fields for the same customer extracted from multiple sources.

Analysis: The act of evaluating the data retrieved from the data warehouse.

Analytics applications: Processes that produce information for management decisions, usually involving demographic analysis, trend analysis, pattern recognition, drill-down analysis and profiling.

Examples of analytics applications include: customer segmentation, customer probability models, campaign measurement, up-sell opportunities, cross-channel analysis, sales distribution analysis, cross-sell opportunities, trigger inventory analysis, supply chain analysis, customer quality analysis, channel satisfaction measurement, click stream analysis, backlog analysis, churn analysis, interaction analysis, booking analysis, billing analysis, distribution analysis, retention analysis, delivery analysis, fulfillment analysis, and promotion effectiveness.

Anomaly: A deviation, irregularity, or an unexpected result. A data anomaly may occur when a data field defined for one purpose is used for another. Examples of anomalies are negative numeric fields that should be positive (negative number of dependents), abnormally high numeric values (person weighing 3000 pounds), pairs of values in related columns that make no sense (male patient having a hysterectomy).

Architect: A person or team who defines how the environment for the data warehouse, analytics application, or operational system is built.

Architecture: A framework for organizing the planning and implementation of data resources. The set of data, processes, and technologies that an enterprise has selected for the creation and operation of information systems. The blueprint that describes the environment that the data warehouse, analysis application, or operational system is built.

ASP – Application Service Provider: A company whose business is providing application services for its client companies. Such applications can include both tactical systems, such as billing systems, or strategic solutions such as CRM. (CRM ASPs currently account for over half the ASP market.)

Atomic data: Data at its most granular and detailed level.

Attributes: In logical data modeling, attributes of an entity refer to the properties of that entity. Each property will have one distinct value per instance of the entity. Example: Entity = Automobile, Attribute = color, attribute value = red. When logical models are translated into physical data models, entities become tables, and attributes become columns. Note: there is not necessarily a 1:1 correlation between the logical model objects and the physical model objects.

Availability: The percentage of time during scheduled hours that the system can be used. It also can refer to the days/week and the hours/day that the system is scheduled for use. See Service Level Agreements

Back-end: Populating the data warehouse with data from operational source systems.

Base table: In relational databases, tables are defined as temporary or base. Base tables are the tables that are created by the CREATE TABLE command and are used for persistent storage.

Batch windows: The time that is required to run the ETL process from beginning to end.

Best-of-breed: Refers to the most effective, powerful, functional and optimal choice of product in each category of tool. As organizations choose tools, they must decide whether they wish to chose a suite of products from the same vendor (where some of the tools in the suite are not terrific) or choose the best product in each category, i.e., best of breed, and integrate those tools themselves.

Best practices: Processes and activities that have been shown in practice to be the most effective.

Beta release: A version of the vendor's software that is given to selected installations prior to the product becoming generally available. This version is often not free of defects.

Big Bang (approach): Delivering all the intended functions of the data

warehouse at the same time.

Bitmapped indexes: This is an alternate (to B-tree) indexing mechanism that involves building streams of bits where each bit is related to the column value for a single row of data in a table. The use of bitmapped indexes on low-cardinality fields (fields that have few possible distinct values) improves query performance significantly.

Boilerplate: Standard verbiage that can be used multiple times for the same purpose. Vendors respond to RFPs with boilerplate so they do not have to write the same material multiple times.

Business analyst: The person whose job it is to analyze the operation and data of the business to develop a business solution.

Business drivers: The tasks, the information and the people that promote and support the goals of the enterprise. The requirements that describe what the business wants (e.g., more quality data, faster response to queries). A problem in the business that is important enough to spell the difference between success and failure for an organization.

Business intelligence (BI): Normally describes the result of in-depth analysis of detailed business data. Includes database and application technologies, as well as analysis practices. Sometimes used synonymously with "decision support," though business intelligence is technically much broader, potentially encompassing knowledge management, enterprise resource planning, and data mining, among other practices.
Business process engineering: The analysis and re-design of business processes and associated technology systems, with the goal to eliminate or reduce redundancy and streamline interactions.

Business rules: Policies by which a business is run. The business rules contain constraints on the behavior of the business. The assertions that define data (e.g., the state code business rule might be the 50 United States, the District of Columbia and the U.S. Territories) from a business point of view.

Business sponsor: Manger or executive who acts as visionary for the data warehouse program and can articulate how the data warehouse can drive business improvements. Establishes the "need, pain, or problem" the data warehouse will solve, serves as a tiebreaker for issues during the project, and might actually fund some or all of the data warehouse development.
See Sponsor

Business timestamp: A business timestamp is a timestamp that is generated by a business event and not a result of a systems operation. Examples are: sales_timestamp, order_timestamp, shipment_date, etc. Typically, all facts in a Data Warehouse have at least one business timestamp, which can be traced to a transaction in the source operational system.

Business users: Personnel reporting to the line-of-business who access the

data warehouse by writing reports and queries or who use the reports and queries generated by others.
See Users

Caching: As related to caching reports, this involves storing the results of pre-run reports in tables (instead of caching to memory as the usage of the word implies) so that when the user accesses the report for the first time, it seems to run instantaneously. This is a feature provided by the server component of many of the popular OLAP tools.

Campaign Analysis: Campaign analysis provides a measurement of responsiveness to campaigns by households and by individual customers. It provides the ability to measure the effectiveness of individual campaigns and different media and offers the ability to conduct cost-benefit analysis of campaigns.

CEO: Chief Executive Officer

CFO: Chief Financial Officer

Champion: The (high level) person in the organization who supports and promotes the data warehouse, its use, and those who developed and maintain it. A person with sufficient clout in the organization who believes in and sells the idea of the data warehouse and helps solve problems between groups.

Channels: The method/means by which a product or service is marketed, ordered, and delivered.

Charge back: The process of assessing and assigning the costs of a system to the departments that use it.

Check totals: "Check totals" is a loose term used to describe the total sum of the values in an additive column of data across all rows of data that are within scope. This total is usually calculated before and after moving data across platforms or processing data in order to ensure no data was lost.

CIO: Chief Information Officer

Class: A collection of objects that share common properties, common definitions and common behaviors.

Clickstream: Series of page visits and associated clicks executed by a Web site visitor when navigating through the site. Analysis of clickstream data can help a company understand which products, Web site content, or screens were of most interest to a given customer.

CMM: Capability Maturity Model: Developed by the Software Engineering Institute (SEI), the CMM is a representation of the goals, methods, and practices needed for the industrial practice of software engineering. The goal of the model is to have processes that are repeatable, defined, managed, and

optimized.

Conformed dimensions: A dimension defines the organization of the measures (facts), or it is an entity "how" an organization measures a fact. A conformed dimension is a dimension that is agreed upon its use and semantics across the enterprise, which makes it "conformed".

COO – Chief Operating Officer

Consultant: A consultant is someone who provides expertise and can be an advisor or a deliverer of tasks. Consultants are hired for their expertise when the company has none
Consultants often help define the data warehouse strategy and assess the organization's ability to implement the data warehouse.

Contractor: A contractor is a person who provides the delivery of tasks. The contractor might be responsible for building the ETL process or for overseeing the DBA functions. Contractors are hired when the company has a shortage of skilled workers. The company tells them what needs to be done, and the contractors perform the work.

Control totals: The addition of values of specific fields to verify that the ETL job streams have executed properly. Cross footing of numbers to verify that a process (e.g. ETL) has executed successfully.

Corporate information factory (CIF): The framework that exists that surrounds the data warehouse; typically contains an ODS, a data warehouse, data marts, DSS applications, exploration warehouses, data mining warehouses, alternate storage, and so forth.

Cost/benefit analysis: The process by which the value of a project is estimated based on the expected costs compared to the tangible benefits usually expressed as increased revenue, or reduced cost.

Critical success factor: An element that contributes to the success of a project, without which the project will fail.

CRM – Customer relationship management: Infrastructure that enables delineation of and increase in customer value and the correct means by which to increase customer value and motivate valuable customers to remain loyal – indeed, to buy again. A collection of integrated applications, which facilitate the seamless coordination between the back office systems, the front office systems, and the web. The DSS expansion of CRM Analytics refers to customer-centric analytics applications.

Cross organizational: Includes multiple departments within an organization. A non-redundant and horizontally cross-functional view of the business.

Cross Selling: Selling an additional category of products as a result of the customer's original purchase.

CTO: Chief Technology Officer

Customer Segmentation: Separating customers by factors such as age, gender, educational background, and liking or disliking Wayne Newton.

DA – Data administrator: The role responsible for the enterprise's data resources and for the administration, control, and coordination of all data related analysis activities. The DA has the responsibility for planning and defining the conceptual framework for the overall data environment. The functions of the DA typically include requirements definition, logical data modeling, data definitions, logical to physical mapping, maintenance of inventory of the current system, data analysis, and the meta data repository.

DASD: Rotating magnetic disk storage.

Data architecture: The framework for organizing the planning and implementation of data resources. The set of data, processes, and technologies that an enterprise has selected for the creation and operation of information systems.

Data analysis: The systematic study of data so that its meaning, structure, relationships, origins, etc. are understood.

DBA – Database administrator: The Database Administrator is responsible for the physical aspect of the data warehouse. This includes physical design, performance, and maintenance activities including backup and recovery

Data loading: The process of populating a data warehouse. It may be accomplished by utilities, user-written programs, or specialized software from independent vendors.

Data mapping: The process of identifying a source data element for each data element in the target environment.

Data mart: An implementation of an analytics application serving a single department, subject area, or limited part of the organization. Usually refers to a physical platform on which summarized data is stored for decision support. Data marts are commonly used for specific analysis purposes by a single organization or user group.

Data mining: Discovery mode of data analysis, or analyzing detail data to unearth unsuspected or unknown relationships, patterns and associations that might be of value to the organization. Advanced analysis used to determine certain patterns within data. Most often associated with predictive analysis. A process of analyzing large amounts of data to identify patterns, trends, activities, and content of data content relationships.

Data ownership: Responsibility for determining the required quality of the data, for establishing security and privacy for the data and determining the

availability and performance requirements for the data. Data originators who have the authority, accountability, and responsibility to create and enforce organizational rules and policies for business data.
See Ownership

Data stewardship: Responsibility for the quality of the business data; an information expert about a particular subject area.
See Stewardship

Data Warehouse Manager: The data warehouse has overall responsibility for all the organization's data warehouse initiatives, for data warehouse standards, and for data warehouse tools. The data warehouse project managers may report to the data warehouse manager or they may report to individual sponsors.

Data Warehouse Project Manager: See Project Manager

Data quality: The degree of excellence of data. Factors contributing to data quality include: the data is stored according to their data types, the data is consistent, the data is not redundant, the data follows business rules, the data corresponds to established domains, the data is timely, the data is well understood, the data satisfy the needs of the business, the user is satisfied with the validity of the data and the information derived from that data, the data is complete, and there are no duplicate records. For example, this means that a customer's name is spelled correctly and the address is correct.

Data staging: The storage of data prior to it being loaded into a data warehouse or data mart.
See Staging area

Data Warehouse: A collection of integrated, subject-oriented databases designed to support the DSS function, where each unit of data is relevant to some moment in time. The data warehouse contains atomic data and lightly summarized data.

DDL – Data definition language: The SQL syntax used to define the way the database is physically organized.

Deadline: The point in time by which a project must be completed.

Deliverable: The tangible output from a task or a project, e.g. logical model, project agreement, database design or application.

Delta: A change, e.g. the difference from one period to the next.

Demo: Short for demonstration as in a vendor demonstration of software to impress the users.

Denormalization: Data or data design elements that do not conform to the rules of data normalization. Denormalized data structures are often used in

databases to provide rapid access for specific user needs. Denormalization usually results in some degree of data redundancy in a data record. A process of combining like data into a single entity (table or file). This combining will create duplicate data.

Departmental systems: A data mart implementation that serves the needs of only a single department such as Human Resources or Finance
See enterprise systems

Derivations: The transformation of data in the ETL process in which the data is created through the use of an algorithm based upon data from multiple sources.

Derived data: A new data element that is created from or composed of other data elements.

Design review: A peer review of project deliverables, such as design specifications, program code or test specifications. The objective of the review is to find weaknesses, errors and problems. The process where different groups are given access to the design to provide input on how it might be changed to 1) work best with the tools selected or 2) be complete in its solving the problem.

Dimensional hierarchy: A dimensional hierarchy refers to the different levels of data within a dimension that data can be rolled up to or down to for analysis. This can be represented in a data model by a series of related tables with parent-child relationships (snow-flaked schema's) or by multiple columns within a dimension table (standard star schemas) called hierarchy columns. Example: the dimensional hierarchy of a sales organization could include the following levels: salesperson, branch, territory, region, company.

Dimension data: An entity used to describe, qualify, or otherwise add meaning to "facts" in a star schema fact table. Dimensions are the "by" items in analysis of facts "by" product, market, time, period, etc. Descriptive data that describes the measurements (facts) that business users wish to analyze.

Domain (synonym valid values): A set of data values which represent the full range of allowable values that may be used for a given data attribute. Defines validity criteria for a particular column or field. Domains include data types and valid values. For example, Gender could be a domain defined as have the data type of Character of 1 byte containing "F" for Female, "M" for Male, and "N" for Not so Sure.

DSS: Decision Support System
See Decision Support System

EIS: Executive Information System: A system that lets upper management view the organization's performance at a highly summarized level and usually in a graphical representation.

End-user: See User, Business User

Enterprise data model: A logical data model that incorporates all the important components of an enterprise data architecture. Components include entities, attributes, relationships, rules and definitions stated in business terms. A schematic defining the data and their relationships that is applied to the whole organization. Diagram of a single non-redundant view of business data, showing how data is used by the business activities of an organization.

Enterprise Data Warehouse: A collection of data that can be defined and shared across the whole enterprise along the lines of common dimensions to be used for analysis.

Enterprise systems: Systems that support and are used by the entire enterprise
See departmental systems

Entity: A person, place, thing, concept or even about which an organization collects data.

ERP: Enterprise Resource Planning: Tying together and automating of diverse components of a company's operations, including ordering, fulfillment, staffing, and accounting. This integration is usually done using ERP software tools.

ETL: Extract/Transform/Load: This is the process of extracting data from their operational data sources or external data sources, transforming the data which includes cleansing, aggregation, summarization, integration, as well as basic transformation (1 becomes "Male" 2 becomes "Female"), and loading the data into some form of the data warehouse (ODS, enterprise data warehouse, data mart). ETL can also refer to the vendor software that performs these processes.

FAQs: Frequently Asked Questions: Questions that are repeated, usually asked by the users of the help desk or of the project support team. Software vendors also have FAQs which are usually asked by technical people who support the vendors' software. To minimize support requirements and to assure a consistent response, FAQs are normally captured, validated, and made available through a web site.

Fact table: The central table in a star join schema, characterized by a composite key, each of whose elements is a foreign key drawn from a dimension table. Facts are information about the business, typically numeric and additive. A table that contains the measures that the business users wish to analyze to find new trends or to understand the success or failure of the organization.

Federated database system (FDS): A federated database system is a collection of independently managed, heterogeneous database systems that allow partial and controlled sharing of data without affecting existing applications. An FDS presents an enterprise view of data.

Foreign keys: Foreign keys are columns on one table that are inherited from the primary key of another table by means of a dependent or independent relationship.

Front-end: The access and analysis piece of the data warehouse architecture.

FTE: Full time employee, Full time equivalent

FTP – File transfer program: A program that transfers data from one computer to another.

Gap Analysis: The difference between what is needed and what is available. The difference between where you are and where you want to be.

Global 2000: The 2000 largest companies worldwide.

Goal: An objective to be achieved within a specific period of time.

Granularity: The level of the measures within a fact table represented by the lowest level of the dimensions.

Hard dollar (benefits): Tangible benefits that can be measured. Hard dollar benefits can result from an increase in revenue or a reduction in cost.

Historical data: Data from previous time periods, in contrast to current data. Historical data is used for trend analysis and for comparisons to previous periods.

Infrastructure: The architectural elements, organizational support, corporate standards, methodology, data, processes, and physical hardware/network, etc. that make up the data warehouse environment.

Integration: The activity of combining data from multiple data sources to present a single collection of data to the warehouse.

Islands of automation: Systems that were developed without consideration for their ability to interface with each other. As a result, data stored in these systems is often redundant and inconsistent.
See silos, stovepipes

IPO: Initial public offering

IT – Information Technology: The department that builds and maintains computer systems.

Iteration: The division of a project in which functionality is provided to the users in a series of phases.

Joins: Within the context of SQL, joining refers to the comparison of similarly valued keys across multiple tables for the purpose of selecting rows of data

from multiple tables. This is done by means of an SQL SELECT statement where the comparison of the keys is performed in the WHERE clause.

Justification: The process by which each project is evaluated to determine if there is financial viability in its implementation. The justification process also allows management to prioritize projects.
See cost/benefit, ROI

Knowledge Transfer: The act of transferring knowledge from one individual to another by means of mentoring, training, documentation, and other collaboration.

Legacy system: Any existing production or operational system. Legacy systems often provide the source data for the data warehouse.
See Operational Systems

Libraries (queries and reports): Sets of programs that have been created, fully tested, quality assured, documented, and made available to the user community. The programs in these libraries are variously called canned, predefined, parameterized, or skeleton queries/reports. They are launched by the user, who only enters a variable such as a date, region number, range of activity or some other set or sets of values the program needs to generate a query or report.

Line of business: Divisions of a company responsible for the production and creation of the organization's products and/or services. IT, HR and Accounting are not lines of business.

Logical data model: An abstract formal representation of the categories of data and their relationships in the form of a diagram, such as an entity-relationship diagram. A logical data model is process independent, which means that it is fully normalized, and therefore does not represent a process dependent (e.g. access-path) database schema.

Market Penetration: The percentage of the market owned by a company as represented by share of revenue.

Matrix management: A reporting structure in which the manager does not hold the performance and payroll card of the subordinate. This is synonymous with dotted line responsibility.

Mentor: A person who provides guidance and recommendations to a more junior person for courses of action and behavior.

Meta data: "Data about data." Usually refers to agreed-on definitions and business rules stored in a centralized repository so business users – even those across departments and systems – use common terminology for key business terms. Can include information about data's currency, ownership, source system, derivation (e.g. profit = revenues minus costs), or usage rules. Prevents data misinterpretation and poor decision making due to sketchy

understanding of the true meaning and use of corporate data.

Methodology: Proven processes followed in planning, defining, analyzing, designing, building, testing, and implementing a system.

Metrics: Any type of measurement. Metrics could include business results, quantification of system usage, average response time, benefits achieved, etc. The measures that an organization believes is vital for its success.

Milestone: A tangible event used to measure the status of the project. Markers during the execution of a project that shows the movement of a project in the right direction.

Mission: A high level set of goals of the organization. For example to be the low cost producer or the company with the highest level of customer satisfaction.

MPP – Massively Parallel Processing: A parallel hardware organization that de-emphasizes the sharing of memory resources.

Multidimensional: The aggregation of data along the lines of the dimensions of the business, e.g. sales by region by product by time.

Near-line storage: Data storage that is not on-line and not with immediate access.

Networking:
(1) Connecting with people of like interests for the purpose of uncovering opportunities, identifying landmines and learning of best practices.

(2) The ability to tie more than one component together through protocols (e.g. TCP/IP)

Object: An instance which is a member of a class.

Objective: Desired outcome of the delivery of the project. An objective can be measured.

OCM: Organizational Change Management

OLAP – Online Analytical Processing: "Drilling down" on various data dimensions to gain a more detailed view of the data. For instance, a user might begin by looking at North American sales and then drill down on regional sales, then sales by state, and then sales by major metro area. Enables a user to view different perspectives of the same data to facilitate decision-making.

OLTP – Online transaction processing: Defines the transaction processing that supports the daily business operations.

OO – Object oriented: A self-contained module of data and its associated

code.

Operational data: Data that supports the productions systems that run the business. This includes, but is not limited to, OLTP systems.

Operational system: The system that creates, updates and accesses production systems. They do not access, or update decision support systems. See Legacy system

Organizational change management: Major change is defined as those situations in which performance of job functions require most people throughout the organization to learn new behaviors and skills. Major change encompasses an entire workforce and can focus on innovation and skill development of people.

To some degree, the downside effects of change are inevitable. Whenever groups of people are forced to adjust to shifting conditions, discomfort will occur. The key is to proactively recognize the effects of change, plan for the change, and develop skill sets and tools to support the change and inevitable discomfort associated with it. Without this proactive approach, the risk of poor project implementation increases significantly and reduces the opportunity to achieve expected compliance.

Outsourcing: Assigning responsibility for all or a portion of the activity and tasks involved in developing and/or running and maintaining a system to a vendor outside of the organization.

Ownership, Owners of source data: One of the more controversial and disputed ideas. The person or group who has responsibility for determining who can access the data warehouse (security), the domains of the data, the performance and availability requirements.
See Data Ownership

Pain: An unfulfilled business need that jeopardizes the success of the organization.

Parallelism: The ability to run the same process simultaneously (in parallel) within more than one processors.

Partitioning: The ability to divide a table into pieces (partitions). The division can be horizontal (by data value – for example by date) or vertical (by columns – for example, most used columns in one partition, the least used columns in another partition.).

Periodicity: The frequency of load/update/refresh of the data warehouse, e.g. daily, weekly, monthly.

PERT Chart: A graphical representation showing the critical path for a project applied to a calendar.

Phasing: The method of delivering the data warehouse in separate groupings of functionality to particular groups of users rather than delivering everything all at once to all the intended users.

Physical Data Model: A formal representation of data and their relationships in the form of a diagram, depicting the physical placement of data in a database. A physical data model is process dependent, which means that it is denormalized to provide maximum performance efficiency. It is commonly referred to as logical database design or database design schema.

Pilot: The initial implementation of a data warehouse. A pilot is always a subset of the intended function and would include a subset of the total set of users. A partially built system to show the capabilities of a full implemented system. A pilot should not become a live system, but usually does. A pilot, proof of concept and prototype are sometimes used synonymously.

Platform: The hardware, operating system and database management or file system
on which the data warehouse runs.

Political agenda: The plans of an individual to enhance his or her position in the organization.

Power users: Knowledge workers who are capable of writing complex queries and reports with little need for help.

Primary key: Refers to the column(s) on a relational table that uniquely define a row of data on that table.

Project agreement: A document outlining the scope of a project including the deliverables, the functions, tools to be used, service level agreements, responsibilities and schedule. The project agreement sometimes includes the anticipated milestones.

Project Manager: Sometimes referred to as the data warehouse project manager, the Project Manager has overall responsibility for a project's successful implementation. The Project Manager defines, plans, schedules, and controls the project. The project plan must include tasks, deliverables and resources – the people who will perform the tasks. The manager will monitor and coordinate the activities of the team, and will review their deliverables. If contractors and consultants are used, the Project Manager assigns the tasks, monitors activities and deliverables and assures that knowledge transfer is indeed taking place.

Project Management Office: Sometimes called project office. This is the office or department responsible for establishing, maintaining and enforcing project management processes, procedures, and standards. It provides services, support, and certification for project managers.

Proof-of-concept: Software trial that allows a prospect to try out the product

before buying it. Delivers a realistic slice of functionality and is often used as the foundation for the first application. A quickly built system to show the capabilities of an idea. A proof-of-concept should not become a live system, but usually does. A pilot, proof of concept and prototype are sometimes used synonymously.

Prototype: A less formal experimental and experiential development process of a proposed application for the purpose of demonstrating some or all of its functional capabilities. A prototype does not have the same rigorous testing, documentation, and implementation requirements as a software release or an application does, and should therefore never be implemented as-is.

Quality: The absence of any defect. The characteristics of a system that conforms to the original design. A system of quality would have the following characteristics: 1. Maintainability (easy to add new functions), 2. Conformance to specifications (fulfilling end user requirements), 3. Long mean time to failure (few bugs and abnormal terminations), 4. Performance that is adequate or as expected, 5. Well tested for functionality, user interface, and performance, 6. Well documented, 7. Easy to use, and 8. Uses standard interfaces.

QA - Quality Assurance: The department, role or process responsible for validating that which is proposed to ensure a correct outcome. The planned and systematic activities to provide confidence that a product or service will fulfill requirements for quality.

RAD: Rapid Application Development
A process where the time is set (timeboxed) and a small set of deliverables is implemented in a reasonably short period of time.

RDBMS – Relational database management system: e.g. DB2, Oracle, SQL Server, Sybase

Real time: Data that is captured, and made available as it is happening. Real time data reflects the latest status of the organization's operational transaction data. Current moment in time. Real time refers to what is happening to any piece of data right now. For analysis, some people want to see current rather than historical data as is the case with most data warehouses.

Recursive: A relationship between two instances of the same entity, as in "recursive data design".

Referential integrity: The concept of enforced relationships between tables based on the definition of a primary key and foreign key.

Release concept: A new approach to development that produces a fully tested, fully documented, high-quality, but only partially functioning application until the final release, which completes the application. The release concept severs the notion that a project deliverable must equal a complete application. Instead it tightens and expands on the concept of a pilot by producing a partially functioning application, which is refined and enhanced several more

times through several more releases before it becomes a fully functioning application. This concept is the embodiment of iterative development and is fully compatible with XP (extreme programming) and the new agile and adaptive methodologies.

Resources: People and budget needed to perform the data warehouse tasks

RFP - Request for Proposal: A formal request to a vendor to submit a proposal to provide a product or service.

ROI – return on investment: Usually represented as a percentage of tangible monetary value in relation to the cost of the system.

Rolled up: Aggregated to a higher level

Scalable: Ability to increase the number of users, the size of the databases and the complexity of the queries and reports without having to replace the existing platform or architecture.

Scope: An itemized accounting and definition of the agreed upon project deliverable in terms of functionality as well as data. In data warehousing, the data scope is more critical than the functional scope for correctly estimating the development effort.

Scope creep: The addition of new requirements, source data or users to the initial agreement of what the project will be delivering.

Semantic layer: A layer between the end-user tool and the database. This allows the end-user tool to present the data most effectively for the end-user understanding and then to generate the proper query to the database.

Service level agreement (SLA): The definition of a level of service provided by the IT department for a particular system. Service level agreements can be established for availability (24 hours/day, 7 days/week and 98% during scheduled hours), for performance (response time for 95% queries in 1 minute or less), for timeliness of the data (weekly data available 6 AM Monday morning), or for other reasons. Contract with a service provider – be it an internal IT organization, an ASP, or an outsourcer – specifying discrete reliability and availability requirements for a given system. Might also include such requirements as support of certain technology standards or data volumes. Outsourcer's failure to adhere to the terms laid out in the SLA could result in financial penalties.

Sign-off: The process of agreeing – in writing – to the scope of a project or the acceptability of a deliverable.

Silo, siloized: A silo system cannot easily integrate with any other system. This means we have multiple versions of the same data, violating the idea of a single version of the truth.
See stovepipe and Island of automation

Single version of the truth: A primary goal of the data warehouse wherein the data to be accessed resides in only one database so that there will be no conflicting data and no inconsistent reports.

Shelfware: Software that is not being used as in "sitting on the shelf".

SMP – Symmetrical Multi-Processing: A parallel hardware organization that emphasizes the sharing of memory resources.

Snowflake structure: Snowflake is a star schema with normalized dimensions.

Source data: The data from the operational or legacy systems that feed the ETL process.

Source system: An operational system, or ODS that is used as the source or input to the ETL process.

Sponsor: The person in the organization, usually from the business side – who supports the project. This person should be someone with power, money and commitment to the project.
See Business Sponsor

Staging area: A staging area is where the ETL programs execute and where the source data is prepared for the data warehouse.
See Data Staging

Stakeholders: People who have a vested interest in the success of the project or are involved in the implementation of the project.

Standards: A standard is "Thou shall" while a guideline is a recommendation, more like "You should if your situation warrants." Data warehouse standards examples include: meta data, terminology, data stewardship, and privacy.

Star schema: A modeling paradigm that has single object in the middle (fact table) connected to a number of objects (dimensions tables) around it radially.

Stovepipe: A stovepipe system cannot easily integrate with any other system. This means we have multiple versions of the same data, violating the idea of a single version of the truth.
See silo and islands of automation

Strategy: Approach taken that will affect the overall direction of the organization and will establish the organization's future environment.

Subject areas: Data Subject Area: Fundamental entities that make up the major components of the business, e.g. customer, product, employee.

Function Subject Area: A business function or business activity, e.g. sales, order processing, inventory.

Suite (of products): A collection of software products from the same vendor – either developed or bundled by that vendor. The idea is to provide a complete set of tools from modeling through to access and analysis. Range of functional software modules that interact with each other. Suites should eliminate integration complexity.

Supply chain: The management of the components, manufacturing and distribution of a manufactured commodity. The supply chain management includes warehousing and tracking inventory.

Systems integration: The art and science of integrating processes, functions, people and data so the end result is a seamless and tight knit system.

System timestamp: A system timestamp is a timestamp that is generated by a systems operation. Examples are : record_create_date, last_update_date,...

SWAT team: A small team of skilled and experienced practitioners who can pull a failing project out of the ditch. This team does not tolerate political interference as it makes decisions and takes actions to bring the project to fruition.

Tactical: Approach taken to achieve a specific objectives or to solve a specific problem.

Target: The database into which data will be loaded from a source database or file; the data store that is accessed by the users.

Terabyte: 1000 Gigabytes.

Third normal form: A database in which each attribute in the relationship is a fact about a key, the whole key and nothing but the key. Usually refers to a fully normalized structure.

Tie (and foot): The process of validating the number of rows, summarizations, and monetary totals of the source data to the data loaded into the data warehouse.

Timely: Data is valuable and useful to analysts only if it represents organizational activities that are reasonably current. Timeliness is a function of the users' requirements for currency and is consistent with user expectations. Timeliness is usually measured by how soon the data is available after some distinctive end-of-period such as "two days after the close of the month." The act of getting the data to the users at the most opportune time.

Time dimension: A table of descriptive attributes about the date/timestamp, e.g. Day of week, Month, Quarter, Season, Year, Century, Holiday, etc.

Time variance: A characteristic of a data warehouse that defines the moment in time that the data or variant of the data is valid. If Order No. 123 has a value of $1,500.00 on Dec 1 and $1,700 on Dec 10, Dec 1 and Dec 10 shows us the

time variance of Order No. 123.

Topology: The manner in which the components of a subject are arranged or interrelated.

Total Cost of Ownership: The cost to the organization for the initial implementation and the maintenance of the system.

Transformation: The manipulation of data to bring it into conformance with the business rules, domain rules, integrity rules, and with other data within the warehouse environment.

Triage: The process by which projects or activities are prioritized to determine which should be attempted first, second, etc. and which projects or activities should never be done at all. This process applies to the cleansing process to determine which data should be cleaned first, second, etc. and which data should not be cleaned at all. Triage considers the value of cleansing, the complexity and the cost and the order in which the cleansing should be accomplished.

Trickle feed: The process by which data updates the target database a little at a time. This is in contrast to massive updates that take place after the close of a period such as the day, month or quarter. The process of feeding data from one system to another in either real-time or small time intervals.

UPC – Universal product code: A unique bar code embossed on every product used for inventory control.

User: A knowledge worker, a business analyst, a statistician, or a business executive who will access the data in the data warehouse to perform some type of business analysis.
See Business Users

Value added: The notion of additional benefit being provided by some activity or service.

Virtual enterprise data warehouse: An enterprise data warehouse constructed of multiple data marts and a request broker computer application. The data warehouse does not physically exist except through out the formation of the integrated data marts.

Vision: The direction of the data warehouse – what it is intended to accomplish.

Visionary: The person in the organization who articulates the data warehouse direction – what it is intended to accomplish.

Visualization: The presentation of results in a format other than just numbers with a display that may include graphs, and charts making copious use of colors and figures.

VLDB (Very Large Database): The perception of what constitutes a VLDB continues to grow. A one terabyte database would normally be considered to be a VLDB.

Work Breakdown Structure: A detailed list of tasks to be performed on the project.

Workload: The quantity of processing to include the machine cycles and the disk I/Os.